HH-FRBC: Halving Hierarchical Fuzzy Rule-Based Classifier

Azam Amouzadi a,∗
Abdolreza Mirzaei a

aElectrical and Computer Engineering Department, Isfahan University of Technology, Isfahan, Iran.

ARTICLE INFO.
Article history:
Received: 14 May 2013
Revised: 04 May 2014
Accepted: 1 October 2014
Published Online: 28 February 2015

Keywords:
Mamdani Fuzzy Rule-based Classification Systems; Hierarchical Fuzzy Rules; Fuzzy Entropy

ABSTRACT
The main objective of this article is to improve the accuracy of Mamdani fuzzy rule-based classification systems. Although these systems tend to perform successfully with respect to interpretability, they suffer from rigid pattern space partitioning. Therefore, a new hierarchical fuzzy rule-based classifier based on binary-tree decomposition is proposed here to develop a more flexible pattern space partitioning. The decomposition process is controlled by fuzzy entropy of each partition. Final rule sets obtained by this proposed method are pruned to overcome the over fitting problem. The performance of this method is compared with some fuzzy and non-fuzzy classification methods on a set of bench mark classification tasks. The experimental results indicate a good performance of the proposed algorithm.

© 2014 JComSec. All rights reserved.

1 Introduction
Fuzzy rule-based systems (FRBSs) are valuable computational intelligence based tools, used for system-modeling [1]. The main interest in using FRBSs arises from the fact that these systems allow us to encounter with the imprecise, noisy, or even incomplete information, which usually exists in most of the real world problems. In addition, fuzzy rules are able to describe nonlinear input/output correlation [2, 3].

Working with FRBSs led to the existence of two different types of system modeling: linguistic modeling represented by Mamdani FRBSs and fuzzy modeling represented by Takagi-Sugeno-Kang (TSK) FRBSs. According to the importance of the interpretability or the accuracy, one of the Mamdani or Takagi-Sugeno methods is adopted. The outstanding feature of Takagi-Sugeno system appears clearly when the accuracy is the main goal, whereas Mamdani system should be used when more attention is drawn to interpretability. Since the focus of this article is on higher accuracy, the Mamdani-type fuzzy rule-based systems are used for improving its accuracy for classification. The generic structure of a Mamdani FRBSs comprises of two main components: the knowledge base (KB) which stores the available knowledge about the problem in the form of fuzzy “IF-THEN” rules and a fuzzy reasoning method (FRM), which classifies a new sample with the information given by the KB [4]. The KB includes two different parts: a rule base (RB) and a data base (DB). One of the significant drawbacks related to Linguistic Mamdani FRBSs is the inflexibility of the DB, which imposes hard restrictions to the fuzzy rule structure. This drawback is considered as a loss in accuracy when modeling some complex systems [5]. In fact, if too coarse fuzzy partitioning is used, as a result of developing over general fuzzy rules, the performance of classification system may decrease. In the case that too fine fuzzy partitions are used, many of the created fuzzy subspaces will have no or few training samples. The rules created for such subspaces are too specific and will cause the system to overfit the training sam-
2 Preliminaries

Here, the two components of Mamdani FRBCS, including KB\(^1\) and FRM\(^2\) and fuzzy entropy will be described.

A. The Data base of Mamdani fuzzy rule-based classification system

The data base (DB), contains the linguistic term sets used in the linguistic rules and the membership functions define the linguistic term sets. Each linguistic variable in the problem will be associated with a fuzzy partitioning. An example of a fuzzy partitioning comprised of four triangular-shaped fuzzy membership functions along with their labels is shown in Figure 1.

![Figure 1. Example of a fuzzy partitioning](image)

B. The Rule base of Mamdani fuzzy rule-based classification system

The rule base (RB) contains a set of linguistic rules that are joined together to make a good decision. Each rule is formed by the linguistic term sets defined in DB. Here the following rule is adopted based on the assumptions bellow: It is assumed that there are \(m\) training (i.e., labeled) patterns \(x_p = (x_{p1}, \ldots, x_{pn})\), \(p = 1, 2, \ldots, m\) from \(M\) classes where \(x_{pi}\) is the \(i\)th attribute value \(i = 1, 2, \ldots, n\) of the \(p\)th training pattern. Therefore a typical rule For FRBCS is created as:

\[
\text{Rule } R_j : \text{if } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \text{ then }
\]

\[
\text{Class } = C_k \text{ with } CF = CF_j, \tag{1}
\]

where \(R_j\) is the label of the \(j\)th rule, \(A_{jk}\) is an antecedent fuzzy set, \(C_k\) is the class label, and \(CF_j\) is the certainty factor of \(j\)th rule \([11]\). According to \([12]\), if the antecedent fuzzy sets of the rule \(R_j\) are known, the consequent class and its \(CF\) can be determined through the following four steps:

1. Computing the compatibility grade of each training pattern \(x_p\), with the fuzzy if-then rule \(R_j\), by the product operator as

\[\text{CF}_j = \prod_{i=1}^{n} C_{ji} \text{, where } C_{ji} = \min \left( \mu_{A_{ji}}(x_{pi}), \text{degree of } x_{pi} \text{ in } A_{jk} \right) \]

\[\mu_{A_{ji}}(x_{pi}) = \begin{cases} 1 & \text{if } x_{pi} \text{ is in } A_{ji} \\ 0 & \text{otherwise} \end{cases} \]

\[\text{degree of } x_{pi} \text{ in } A_{jk} = \frac{1}{1 + \frac{k}{d(x_{pi}, \text{center of } A_{jk})}} \]

2. Evaluating the compatibility grade of each training pattern \(x_p\) with respect to the consequent class \(C_k\)

\[\text{degree of } x_p \text{ in } C_k = \sum_{j=1}^{m} \text{CF}_j \]

3. Selecting the class with the highest degree of compatibility

\[\text{Class } = \text{arg max } \text{degree of } x_p \text{ in } C_k \]

4. Calculating the certainty factor \(CF\)

\[CF = \frac{1}{1 + \text{degree of } x_p \text{ in } C_k} \]

5. Knowledge base
6. Fuzzy reasoning method
\[
\mu_{R_j}(x_p) = \mu_{A_{j1}}(x_{p1}) \times \cdots \times \mu_{A_{jn}}(x_{pn}),
\]
where \(\mu_{A_{j1}}()\) is the membership function of the antecedent fuzzy set \(A_{j1}\) and “\(\times\)” is used as T-norm.

(2) Computing the sum of compatibility grades for each class as follows:
\[
\beta_{class\ h}(R_j) = \sum_{x_p=class\ h} \mu_{R_j}(x_p) \quad h = 1, \ldots, M,
\]  

(4) Specifying the certainty factor of the rule \(R_j\) using Equation (6),
\[
CF_j = \beta_{class\ c_j}(R_j) = \frac{\tilde{\beta}}{\sum_{k=1}^{M} \beta_{class\ k}(R_j)},
\]
where
\[
\tilde{\beta} = \sum_{h \neq c_j} \beta_{class\ h}(R_j) / C - 1.
\]

C. Fuzzy reasoning method

A FRM is an inference procedure which classifies a new sample with the information given by the KB. There are different methods for reasoning, which are presented in [10]. In this article the weighted method is used as an inference method. In this method, to classify an input pattern \(x_p\) by \(S\), which is a set of rules, first, the compatibility grade of \(x_p\) with the antecedent part of each rule is calculated through Equation (3). Then the strength of the vote given by each rule is calculated as the product of compatibility grade and certainty factor. After that, the total strength of the vote for each class is calculated as:
\[
\sigma_{class\ h}(x_p) = \left\{ \sum_{R_j \in S, C(R_j)=class\ h} \mu_j(x_p) CF_j \right\},
\]
Finally, the test pattern \(x_p\) will be classified as the class having maximum total strength.

D. Fuzzy Entropy

The fuzzy entropy which is used in this article [13] is based on Shannon’s entropy. If the assumptions made to develop Equation (1) are considered and \(S_{c_j}\) represents a set of elements of class \(j\) on the universal set \(X\) in an interval value of \(R_j\) (it is a subset of the universal set \(X\)), the fuzzy entropy of a fuzzy rule \((R_j)\) will be calculated through the following steps:

- Calculating the match degree \(D_j\) of the elements of class \(j\) with the fuzzy set \(\tilde{A}\) for each rule by Equation (8).
\[
D_j = \sum_{x_p \in S_{c_j}} \mu_{\tilde{A}}(x_p) / \sum_{x_p \in X} \mu_{\tilde{A}}(x_p),
\]
- Calculating the fuzzy entropy \(FE_{(C_j)\tilde{A}}\) for the elements of class \(j\) in an interval of \(R_j\) by Equation (9).
\[
FE_{c_j}(\tilde{A}) = -D_j \log_2 D_j,
\]
- Calculating the fuzzy entropy \(FE_{\tilde{A}}\) for the elements within the \(R_j\)’s interval by Equation (10).
\[
FE_{\tilde{A}} = \sum_{j=1}^{M} FE_{c_j}(\tilde{A}),
\]

3 The Proposed Hierarchical Classifier

One of the drawbacks of Mamdani FRBCS is the lack of accuracy caused by the inflexibility of the concept of linguistic variables [3]. One of the most important reasons of this inflexibility can be attributed to the rigid partitioning of the input-output spaces [14]. In order to improve the accuracy of FRBCS a new hierarchical classifier, named Halving Hierarchical Fuzzy Rule Based Classifier (HH-FRBC) is proposed in this article to make KB more flexible. To serve this purpose a new KB, named Halving Hierarchical Knowledge base (HHKB) is presented. HHKB is composed of Halving Hierarchical Data Base (HHDB) and Halving Hierarchical Rule Base (HHRB).

3.1 HHDB

The linguistic terms of the HHDB are formed in different levels in a hierarchical manner. In each level the DB will be updated using the information of the linguistic terms of the previous level. The number of linguistic terms in the fuzzy partitions of the first level is 2. The membership functions (MFs) used in the first level of HHDB are shown in Figure 2.

Here, the MFs of each new level are formed by halving the interval of the MF in previous level. The process of creating the DB of the second level by means of the DB of the first level is illustrated in Figure 3. The process of creating the DB of the third level by means of the DB of the second level is shown in Figure 4.

If the interval values of the MF of previous level are considered as \(L\) and \(H\), the parameters of the MFs of the new level of HHDB are calculated by Equation (10)–12. Explanations of each Equation follows:
Provided that one of the interval values of a MF which is to be halved equals to one of the interval values of the variable Figure 3, the parameters of the MFs of new level are calculated by Equation (11) or Equation (12). The Equation (11) and Equation (12) are used for halving $A_1$ and $A_2$ respectively.

\[
\begin{align*}
\{ & a_1 = L, \quad b_1 = L + d, \quad c_1 = H - d, \\
& a_2 = L + d, \quad b_2 = H - d, \quad c_2 = H + d, \\
\} 
\end{align*}
\]

(11)

or

\[
\begin{align*}
\{ & a_1 = L - d, \quad b_1 = L + d, \quad c_1 = H - d, \\
& a_2 = L + d, \quad b_2 = H - d, \quad c_2 = H + d, \\
\} 
\end{align*}
\]

(12)

In case the interval values of the MF is considered to be halved would not equal to the interval values of the variable (Figure 4), Equation (13) is applied.

\[
\begin{align*}
\{ & a_1 = L - d, \quad b_1 = L + d, \quad c_1 = H - d, \\
& a_2 = L + d, \quad b_2 = H - d, \quad c_2 = H + d, \\
\} 
\end{align*}
\]

(13)

Where, in these three Eqs.: $d = (H - d)/4$.

Similarly, the MFs of each new level can be constructed by the MFs obtained from the previous level to form the HHDB.

### 3.2 HHRB

The main objective of developing an HHRB is to model the problem space in a more accurate manner based on the data distribution. The HHRB is formed based on the MFs created in different levels of HHDB. All possible fuzzy rules of the new level could be created as follows: For each existing rules of $k^{th}$ level, two fuzzy rules of $(k + 1)^{th}$ level is created by halving the MFs related to one of the variables of $k^{th}$ level rule , and forming the rules corresponding to new MFs.

It is worth mentioning that, for each rule of $k^{th}$ level, the best variable is chosen for halving. All possible fuzzy rules of the first and the second level are shown in Figure 5 and Figure 6 respectively.

**Figure 2.** The membership functions which are used in the first level of HHDB $A_1, A_2$

**Figure 3.** Mapping between terms from first level to second level

**Figure 4.** Mapping between terms from second level to third level

### 3.3 The creation of linguistic values for rules of different levels

In this proposed system, like other FRBCs, the antecedent of fuzzy rules contains linguistic variables.
Figure 5. All possible first-level rules: \( R_1 \) and \( R_2 \) are created by halving the initial space along the \( X_1 \), \( R_3 \) and \( R_4 \) are created by halving the initial space along the \( X_2 \).

Figure 6. All the possible unique second-level rules: (a) Second-level rules which are created by halving \( R_1 \) and \( R_2 \) of the Figure 2 along the \( X_2 \) or \( R_3 \) and \( R_4 \) of the Figure 2 along the \( X_1 \); (b) Second-level rules which are created by halving \( R_1 \) and \( R_2 \) of the Figure 2 along the \( X_1 \); (c) Second-level rules which are created by halving \( R_3 \) and \( R_4 \) of the Figure 2 along the \( X_2 \).

Figure 7. (a) \( R_1 \) is a rule in the third level and \( x_1 \) is chosen in all its three levels to be halved (b), (c) and (d) are the steps of creating the linguistic value corresponding to \( x_1 \) of rule \( R_1 \).
Algorithm 1  The outline of the proposed algorithm

while 1 do
    for each existing fuzzy rule do
        Calculate the Fuzzy Entropy by Equation (9)
    end for
    if there are not any rules whose entropy is more than a threshold then
        Exit
    else
        Step0: Select the rules whose entropy is more than a threshold
        for each selected rule do
            Step1: Compute the Fuzzy Entropy by halving each variable
            Step2: Select the variable that causes the best entropy
            Step3: Exchange the selected rule for two created rules produced by halving it along the variable which causes the best entropy
        end for
    end if
end while
Prune the final rule set by validation set

and values. Here, the linguistic values are obtained based on the rule’s level. For example, consider R1 in Figure 7 (a), which is a rule in the third level. Halving this rule has occurred along the $x_1$ for all three levels. The linguistic value of $x_1$ will be determined as follows:

- After the initial space is halved along $x_1$ (i.e. $A_1$ and $A_2$), the second linguistic value which is equivalent to $A_2$ is selected, Figure 7 (b).
- the subspace $A_2$ of Figure 7 (b) in bold is halved (i.e. $A_{21}$ and $A_{22}$) and the first linguistic value which is equivalent to $A_{2,1}$ is selected Figure 7(c).
- the subspace $A_{21}$ of Figure 7 (c) in bold is halved (i.e. $A_{211}$ and $A_{212}$) and the second linguistic value which is equivalent to $A_{2,1,2}$ is selected, Figure 7 (d).

Consequently, the linguistic value $x_1$ in $R_1$ is represented by $A_{2,1,2}$. Bearing in mind that in this rule, $x_2$ is not selected to be halved, so ”don’t care” is chosen for this variable and $x_2$ is omitted from the rule representation. Therefore, $R_1$ will be represented as follows:

If $x_j = A_{2,1,2}$ then class=$C1$.

3.4 Proposed algorithm

HHFRBC creates rules in a hierarchical manner by means of HHDB and HHRB. The outline of the proposed algorithm is illustrated in Algorithm 1.

It should be mentioned that, the rule that its fuzzy entropy is to be calculated at the beginning of the algorithm is the initial space of the problem.

To show different steps of this algorithm, suppose that the initial space of the problem with some data from two classes is illustrated in Figure 8. For this example the threshold is considered to be 0.

Step0: Since the FE of the initial space is greater than zero, this rule (i.e. the initial space) is selected.

Step1: The FE produced by halving each variable is calculated by Equation (14).

$$FE = \left(\frac{n_1}{N}\right)FE(R_2) + \left(\frac{n_2}{N}\right)FE(R_3), \quad (14)$$

Where, $N$ is the total number of data in $R_1$, $n_1$ and $n_2$ are the number of samples in $R_2$ and $R_3$ respectively. The process of halving $R_1$ along the first and the second variables is illustrated in Figures 9 and 10.

Step2: Since the produced FE through halving the first variable (Figure 9) is less than that of the one produced by halving the second variable (Figure 10), the first variable is chosen as the best one.

Step3: $R_1$ will be replaced with $R_2$ and $R_3$ (Figure 9).

At this point FE will be calculated for $R_2$ and $R_3$ separately using Equation (10) and the step0 to step3 are repeated. This process is iterated until the FE of all rules become zero. After two iterations, the final rule set is shown in Figure 11.

Finally, due to the stopping criterion (i.e. threshold=0 for all rules) used in this article, the final rule set extracted by this newly introduced method may overfit the data. To overcome this overfitting drawback, the final rule set will be pruned using the validation set (i.e. a subset (1/3) of the training data is used as the validation set). To do so, after creating the final classifier, a rule is removed from the rule set, if by removing it, the accuracy of the classifier over the validation set does not decrease.
Experimental Study

In the experiments here, a five-fold cross validation approach (i.e., data is divided into five partitions: one part is used for testing and the remaining is used for training) is applied. For each dataset, the average result of the five partitions is considered as the final result. To detect significant differences among the results obtained in this study, some statistical tests are performed. There are different tests for multiple comparison over multiple datasets [15, 16]. Since there is not any assumption for normality in the nature of our problem, Friedman and Iman and Dovenport tests [17, 18] are chosen to compare the classifiers. Holm method [19] is adopted as a post processing method to compare the best ranking classifier with the remaining ones. A complete description of these tests and the software used for its calculation can be found in [20].

The data sets used in this article will be presented before examining the effect of pruning in the HH-FRBC. In the following stage, a comparison is made between the results of HFRBCS [7] and this proposed structure is made.

Experimental setup and the effect of pruning

Here, 8 data sets from UCI machine learning repository [21] are selected along with their names, number of examples (# Ex), attributes (# Atts) and classes (# Class) (Table 1). Besides, the results of LDWPSO algorithm are obtained by running Keel2 software [20]. Most of the approaches in constructing final rule sets involve greedy heuristics (such as fuzzy entropy reduction) that may overfit the training data and result in low accuracy in the test data. To overcome this drawback, some pre-pruning and post-pruning methods are adopted. A post-pruning method is used here and its effect on accuracy and the number of obtained rules are examined. The results are presented in Table 2. As the results show, the test accuracy is slightly increased after pruning the final rule set and the number of rules is decreased significantly.
Table 1. Data description

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of Attributes</th>
<th>Number of Classes</th>
<th>Number of Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer</td>
<td>9</td>
<td>2</td>
<td>699</td>
</tr>
<tr>
<td>Wine</td>
<td>13</td>
<td>3</td>
<td>178</td>
</tr>
<tr>
<td>Iris</td>
<td>4</td>
<td>3</td>
<td>150</td>
</tr>
<tr>
<td>Vehicle</td>
<td>18</td>
<td>4</td>
<td>846</td>
</tr>
<tr>
<td>Bupa</td>
<td>6</td>
<td>2</td>
<td>345</td>
</tr>
<tr>
<td>yeast</td>
<td>8</td>
<td>10</td>
<td>1,484</td>
</tr>
<tr>
<td>Page-blocks</td>
<td>10</td>
<td>5</td>
<td>548</td>
</tr>
<tr>
<td>Ecoli</td>
<td>7</td>
<td>8</td>
<td>336</td>
</tr>
</tbody>
</table>

Table 2. The accuracy and the number of rules of the proposed method with and without employing pruning method

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Accuracy without Pruning</th>
<th>Accuracy with Pruning</th>
<th>Number of Rules without Pruning</th>
<th>Number of Rules with Pruning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer</td>
<td>0.9617</td>
<td>0.9661</td>
<td>45</td>
<td>7.8</td>
</tr>
<tr>
<td>Iris</td>
<td>0.9466</td>
<td>0.9533</td>
<td>36.2</td>
<td>7.8</td>
</tr>
<tr>
<td>Vehicle</td>
<td>0.6733</td>
<td>0.7124</td>
<td>1895</td>
<td>104.6</td>
</tr>
<tr>
<td>yeast</td>
<td>0.6486</td>
<td>0.6668</td>
<td>2412</td>
<td>262.4</td>
</tr>
<tr>
<td>Bupa</td>
<td>0.7130</td>
<td>0.6927</td>
<td>259.6</td>
<td>48</td>
</tr>
<tr>
<td>Page-blocks</td>
<td>0.9451</td>
<td>0.9425</td>
<td>2802</td>
<td>96.8</td>
</tr>
<tr>
<td>Wine</td>
<td>0.8742</td>
<td>0.8914</td>
<td>25.4</td>
<td>8.2</td>
</tr>
<tr>
<td>Ecoli</td>
<td>0.8358</td>
<td>0.8089</td>
<td>180</td>
<td>31.8</td>
</tr>
<tr>
<td>average</td>
<td>0.8247</td>
<td>0.8292</td>
<td>956.9</td>
<td>70.92</td>
</tr>
</tbody>
</table>

Table 3. Training results of HHFRBC in comparison with other classifiers

<table>
<thead>
<tr>
<th>Dataset</th>
<th>HH-FRBC</th>
<th>Chi3</th>
<th>HFRBCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer</td>
<td>0.9806</td>
<td>0.9806</td>
<td>1</td>
</tr>
<tr>
<td>Iris</td>
<td>0.9933</td>
<td>0.9373</td>
<td>0.9493</td>
</tr>
<tr>
<td>Vehicle</td>
<td>0.9101</td>
<td>0.6587</td>
<td>0.9306</td>
</tr>
<tr>
<td>yeast</td>
<td>0.9871</td>
<td>0.2963</td>
<td>0.6399</td>
</tr>
<tr>
<td>Bupa</td>
<td>0.9884</td>
<td>0.60</td>
<td>0.8424</td>
</tr>
<tr>
<td>Page-blocks</td>
<td>0.9661</td>
<td>0.9260</td>
<td>0.9514</td>
</tr>
<tr>
<td>Wine</td>
<td>0.9943</td>
<td>0.9874</td>
<td>0.9994</td>
</tr>
<tr>
<td>Ecoli</td>
<td>0.9940</td>
<td>0.7833</td>
<td>0.9326</td>
</tr>
<tr>
<td>average</td>
<td>0.9767</td>
<td>0.7712</td>
<td>0.9064</td>
</tr>
</tbody>
</table>

Table 4. Test results of HHFRBC in comparison with other classifiers

<table>
<thead>
<tr>
<th>Dataset</th>
<th>HH-FRBC</th>
<th>Chi3</th>
<th>HFRBCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer</td>
<td>0.9661</td>
<td>0.9121</td>
<td>0.9107</td>
</tr>
<tr>
<td>Iris</td>
<td>0.9533</td>
<td>0.94</td>
<td>0.9267</td>
</tr>
<tr>
<td>Vehicle</td>
<td>0.7124</td>
<td>0.6077</td>
<td>0.6726</td>
</tr>
<tr>
<td>yeast</td>
<td>0.6668</td>
<td>0.2918</td>
<td>0.5748</td>
</tr>
<tr>
<td>Bupa</td>
<td>0.6927</td>
<td>0.5787</td>
<td>0.6259</td>
</tr>
<tr>
<td>Page-blocks</td>
<td>0.9425</td>
<td>0.9161</td>
<td>0.9270</td>
</tr>
<tr>
<td>Wine</td>
<td>0.8914</td>
<td>0.9268</td>
<td>0.9160</td>
</tr>
<tr>
<td>Ecoli</td>
<td>0.8089</td>
<td>0.7833</td>
<td>0.8719</td>
</tr>
<tr>
<td>average</td>
<td>0.8292</td>
<td>0.7445</td>
<td>0.8032</td>
</tr>
</tbody>
</table>

Table 5. Average number of rules for HHFRBC in comparison with other classifiers

<table>
<thead>
<tr>
<th>Dataset</th>
<th>HH-FRBC</th>
<th>Chi3</th>
<th>HFRBCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer</td>
<td>7.8</td>
<td>224</td>
<td>216.3</td>
</tr>
<tr>
<td>Iris</td>
<td>7.8</td>
<td>16.6</td>
<td>18.4</td>
</tr>
<tr>
<td>Vehicle</td>
<td>104.6</td>
<td>227.8</td>
<td>2041</td>
</tr>
<tr>
<td>yeast</td>
<td>262.4</td>
<td>25.4</td>
<td>216</td>
</tr>
<tr>
<td>Bupa</td>
<td>48</td>
<td>43.4</td>
<td>120</td>
</tr>
<tr>
<td>Page-blocks</td>
<td>8.2</td>
<td>120.2</td>
<td>144.9</td>
</tr>
<tr>
<td>Wine</td>
<td>96.8</td>
<td>19.6</td>
<td>39.4</td>
</tr>
<tr>
<td>Ecoli</td>
<td>31.8</td>
<td>43.5</td>
<td>158.9</td>
</tr>
<tr>
<td>average</td>
<td>70.92</td>
<td>90.06</td>
<td>369.36</td>
</tr>
</tbody>
</table>

Table 6. Results of the Friedman and Iman-Davenport tests for comparing performance of the HH-FRBC in all data sets

<table>
<thead>
<tr>
<th>Method</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friedman</td>
<td>0.0075</td>
</tr>
<tr>
<td>Iman-Davenport</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

Total analysis of HH-FRBC

In order to show the advantages of HH-FRBC, its accuracy is compared with two FRBCs namely Chi3 and HFRBCS. The comparison of the results of HH-FRBC and other classifiers on training and test data set are tabulated in Tables 3 and 4 respectively. The number of rules obtained with each method is tabulated in Table 5. The average results in Table 6 indicate that HH-FRBC outperforms other algorithms on training data sets. The results of Table 6 indicate that the highest value of test accuracy is associated with HH-FRBC. The average number of rules of this proposed method is less than Chi3 and HFRBCS (Table 5).

In addition, Friedman and Iman-Davenport tests are used to see if there exists any significant differences in the results. The results of applying these tests are tabulated in Table 6.

Because the P-values produced by Friedman and
Table 7. Results of Holm test. PHH-FRBC is the control method

<table>
<thead>
<tr>
<th>i</th>
<th>Algorithm</th>
<th>P</th>
<th>α/i</th>
<th>Hypothesis (α=0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>HFRBCS</td>
<td>0.11384629800665803</td>
<td>0.025</td>
<td>Not Rejected</td>
</tr>
<tr>
<td>1</td>
<td>Chi3</td>
<td>0.01425529135570033</td>
<td>0.05</td>
<td>Rejected for PHH-FRBC</td>
</tr>
</tbody>
</table>

Iman-Davenport tests in Table 7 are less than 0.05, it could be deduced that there are significant differences among the results obtained by different classifiers. In other words, obtained p-values inform us of the rejection of the null hypothesis (i.e. equality of means). The ranking here is computed through a Friedman test of the 5 algorithms (Figure 12). The value given to each method is acquired by assigning a position to each algorithm with respect to its performance for each data set. The algorithm with the best accuracy on a specific data set receives the first ranking (value 1), next, the algorithm with the second best accuracy receives rank 2 and so on. This assignment is done for all data sets. Finally an average ranking is calculated as the mean value of all ranks. The best ranking method is HH-FRBC (Figure 12). Now the best ranking method (HH-FRBC) is selected as control method and a post-hoc test (i.e. Holm test in this case) is applied to find the algorithms that reject the equality hypothesis. The results of this test is tabulated in Table 7. The corresponding p-values are obtained based on Normal distribution. To show whether the hypothesis is rejected in favor of the control method or not, the obtained P-values can be compared with the associated α/i in the same row of the Table 7. If the p-value is less than associated α/i, it indicates that that, the control method rejects the corresponding hypothesis. According to the above discussion, it is found that, the proposed method rejects Chi3 approach (Table 7).

Although the proposed method achieves a higher ranking than HFRBCS, it does not suffice to reject this hypothesis. Therefore, it may be deduced that both approaches have a similar performances. The obtained results in Table 7 demonstrate the goodness of this proposed hierarchical structure, because HH-FRBC outperforms the simple Chi algorithm in standard classification. Although, HH-FRBS cannot reject HFRBCS, there is a vast difference between the number of rules obtained by these two methods.

5 Conclusion

In this article a new hierarchical fuzzy rule-based classifier named HH-FRBC is proposed based on binary tree decomposition.

Figure 12. Ranking of different classifiers obtained by Friedman test

the decomposition process is controlled by fuzzy entropy of each partition and a pruning method is applied to the final rule set to overcome overfitting problem. The main objective of considering hierarchical structure for rules is generating fine and coarse fuzzy subspaces simultaneously in order to cover the problem space properly. In the experimental study, the adopted method is compared with a simple Chi algorithm in standard classification and the last version of hierarchical fuzzy rule based classification system (i.e. HFRBC). In its statistical sense the performance of this proposed classifier is better than that of the other classifiers on the datasets. It should be added that, the number of rules obtained by the proposed method is less than that of the other fuzzy rule based method. These findings prove the interpretability of this proposed classifier. The essence of the argument is that this proposed method provides trade-off between the interpretability and accuracy.

References


Azam Amouzadi was born in Isfahan, Iran. She received the B.S. degree in computer engineering from Allama Mohaddes Noori Institute of Higher Education, Noor, Mazandaran, Iran in 2006 and the M.Sc. in Artificial Intelligence from Isfahan University of Technology in 2010. Currently, she is the head of computer department in Safahan Institute of Higher Education, Isfahan, Iran. Her research areas include Fuzzy Systems, Machine Learning, Data Mining and Evolutionary Algorithms.

Abdolreza Mirzaei was born in Isfahan, Iran. He received the B.S. (first class honors) degree in computer engineering from Isfahan University in 2001 and the M.Sc. and Ph.D. degrees in Artificial Intelligence from Iran University of Science and Technology and Amirkabir University of Technology, in 2003 and 2009, respectively. He is currently an assistant professor in the electrical and computer engineering department at Isfahan University of Technology, Iran. His research areas include Pattern Recognition, Machine Learning, Data Mining, and Swarm Intelligence.